Network Intelligence Studies
Volume VII, Issue 13 (1/2019)

Andreea DUMITRACHE,
Alexandra NASTU
Bucharest University of Economic Studies

TECHNIQUES OF DATA ANALYSIS. AN APPROACH TO BUSINESS STATISTICS AND ECONOMETRICS

Case Study

Keywords
Statistics;
Time series;
Stationarity;
Normality;
Linearity test;
Data Science

JEL Classification
H20, H50

Abstract

Data can be analysed by using a wide range of techniques. The present article focuses on the use of the basic approaches to statistics and econometrics (descriptive analysis, distribution analysis, logistic regression and tests for model validation) in order to establish a correlation between the individual returns of the securities and a macroeconomic factor. The novelty of the paper consists in designing well-defined steps according to objective criteria of the financial market when evaluating financial assets. Thus, a unifactorial model consisting of several data science techniques is used, which assumes that the profitability of any financial title is in a linear relationship with a macroeconomic variable.The study is based on Apple and market portfolio data series and the results show that there is a strong positive dependence between them.

INTRODUCTION

The company whose data is used in this paper is Apple Inc, a multinational company, based in Cupertino, California. The company designs, develops and markets consumer electronic products, computer software, online services and personal computers. Its most popular hardware products are the Mac line of computers, iPods, iPhone smartphones and iPad tablets. Its online services include iCloud, iTunes store and App store. The current study used Apple asset prices and SP500 market portfolio prices. The data used are daily observations over a period of time between November 27, 2015 and November 25, 2016 (251 observations). These were processed in R software.

The aim is to establish a correlation between the individual returns of the securities and a macroeconomic factor. The paper proposes a new method of evaluating financial assets according to objective criteria of the financial market. Thus, a unifactorial model consisting of several data science techniques is used, which assumes that the profitability of any financial title is in a linear relationship with a macroeconomic variable.

Another aspect worth considering is the volatility of a security, i.e. its sensitivity to market movements. Volatility measures the sensitivity of the stock to market movements. It can be positive (most often) or negative (more rarely) and more or less strong, as market fluctuations accentuate or diminish those of the title. The relationship between the profitability of a security and the profitability of the market is highlighted through the market model (Sharpe, 1964). The market model represents the linear relationship between the individual return on securities and the overall return on the stock market. The function that approximates the correlation between the two returns is a straight line, called a regression line. The most important of the parameters of the regression function is the beta coefficient (volatility coefficient). To effectively test if the model is applicable, one should verify whether the return on the market of a security matches the expected return for it.

According to Sharpe's theory, the security risk is composed of two parts, namely the systematic risk related to the capital market and explained by the dependence on the macroeconomic factor, and the specific risk of each security, which can be removed through diversification.

DISTRIBUTION ANALYSIS

Mean, Coefficient of Correlation and Standard Deviation

During the analysis period, 27.11.2015 and 25.11.2016, the average return on Apple's assets was 0.0002383987, and the average return on the market portfolio was -0.0001894428. The correlation coefficient between the profitability of the Apple asset and the profitability of the market portfolio was 0.583617. It has a significant value, which means that there is a strong link between the return of the Apple asset and the profitability of the market portfolio.

The standard deviation for the profitability of the asset was 0.01535756, and 0.008833525 for the profitability of the market portfolio. The standard deviation shows how much the values are spread over the average. Thus, the profitability of the asset deviated from the average by 0.01, and the profitability of the market portfolio by 0.009.

Asymmetry and Flattening

Kurtosis is a statistical indicator used to describe the degree to which a distribution is flat or peaked (Adler and Adler, 1994). It indicates the degree to which scores cluster in the tails or the peak of a frequency distribution, where the peak is the tallest part of the distribution, and the tails are its ends. The value of the indicator is 6.59 for the return of the asset. In our case it is greater than 3, which means that the profitability of the asset follows a leptokurtic distribution, which is more sharply peaked with heavier tails than a normal distribution, having more values concentrated around the medium tails. At the same time the probabilities for extreme values are high. The value of kurtosis is 4.60 for the profitability of the market portfolio. This is greater than 3, which means that the profitability of the market portfolio follows a leptokurtic distribution.

Skewness is an indicator used in the analysis of a data series distribution to indicate the deviation of the empirical distribution with respect to a symmetrical distribution around the mean. The value of this indicator is 0.349 for the return of the asset. Because this value is positive the distribution is skewed left, having more extreme values to the right. The profitability of the skewness market portfolio is 0.409, so the distribution is skewed left, with more values on the right side.

Quantiles shows how the data set is divided (Angrist and Piche, 2009). Thus, in the case of the profitability of the market portfolio, a quarter of values (25%) is below -0.004730 and 75% are above this level. The median is equal to -0.000155, which means that only 50% of the values are below -0.000155 and 50% are above this value. A quarter of values exceeds 0.00326, while three quarters are below this level. These quartiles also provide us

with information about the maximum and minimum values registered: -0.0241 and 0.03725, respectively. The same can be said about the return on the asset. 25% of the registered values are below -0.00767 and 75% above this value. Only one quarter of the values exceeds 0.00753, while 75% of the values recorded during the analysis period are below this level. At the same time, 50% of the observed values are lower than -0.00044. The maximum value is 0.07032 and the minimum is -0.0610 (Table 1).

The graphic representation of information

Histogram: The graph represented as a histogram in the case of profitability shows that the asset follows a leptokurtic distribution, having more values concentrated around the average, close to 0 (Figure 1).

The histogram of the profitability of the market portfolio follows a leptokurtic distribution, having several values concentrated around the average, close to 0. Most values are in the range [0; 0.005], followed by the interval [-0.005; 0]. No value can be noticed in the interval [0.035; 0.04] (Figure 2).

Graphic representation of the density function: as noticed the market portfolio profitability and the profitability of the Apple asset have the highest density around 0. Also, the probability densities of the two variables have approximately the same trend during the analysed period (Figure 3 and 4).

Boxplot is the way to identify the population according to normality (Adler and Adler, 1994). Figure 5 shows that the minimum value is -0.02 and the maximum is 0.037. It is observed that 25% of the values are below -0.0047, 50% of the values are below -0,0001 and only 25% of the values are above 0.003 (Figure 5).

According to the graph in Figure 6, the minimum value is -0.06, and the maximum value is 0.07. The median is -0.00044, which means that 50% of the values are higher than this and 50% are higher. Also. We note that 25% of the returns on Apple assets are lower than -0.007 and only 25% are greater than 0.007.

The interdependence between the two returns is positive, that is if the market portfolio returns increases, it means that the return on Apple's assets will also increase (Figure 7).

Testing normality with the JarqueBeraTest

The null hypothesis for the JB test is that the distribution is normally distributed (Wu, 1973). As the p-value probabilities for the following JB tests are very small, the distribution of the asset and the distribution of the market portfolio do not follow a normal distribution, as can be seen from the histograms presented in Table2.

Testing stationarity.Dickey-Fuller Test

Next, the correlograms of the two series are analysed to decide whether they are stationary or not (Harvey, 1991).

Figure 8 illustrates that the series of returns of the Apple asset is stationary because no value exceeds the point range. Figure 9 shows that the series of returns of the market portfolio is stationary (the Dickey –Fuller test can be applied as a means of verification).

The series of profitability of the Apple asset and the series of profitability of the market portfolio are shown in Table 3. As the p-value probabilities are lower than the 0.05 threshold, the series are considered stationary.

THE INTERDEPENDENCE BETWEEN THE PROFITABILITY OF THE ASSET AND THE PROFITABILITY OF THE MARKET PORTFOLIO; THE REGRESSION FUNCTIONS

The regression line shows that the asset follows the distribution of the market, because when the profitability of the market portfolio increases or decreases and the profitability of the asset has the same evolution (Figure 10). Between the two returns there is a positive interdependence, which shows how the asset grows when the market grows.

The regression function (Wright, 1921):

$$f(x) = b * x\_1 + a$$

Asset return: b * market portfolio return + a. The value of intercept (a) shows the level of profitability of the asset if the profitability of the market portfolio is 0. The value of the slope coefficient Rm - is different from 0, so this coefficient is statistically significant. It shows the value of the asset when the market grows by one unit. Thus, if the market increases by 1%, then the asset increases by 1.01%.

Adj R-square is 33%, which means that the market variation explains 33% of the asset variation. P-value shows the probability of failing when rejecting the null hypothesis (which says the model is not valid). This value must be very small. In the case of the coefficient Rm the p-value is very small, which means that the probability of error is very low (Table 4).

The confidence interval: with a probability of 95%, according to the graph below the profitability of the market portfolio is within the confidence interval [0.83; 1.19] (Table 5).

VERIFICATION OF VALIDATION CRITERIA

Higher-order autocorrelation
The Breusch–Godfrey Test is applied. H0: the residues are not self-correlated in a higher order. H1: the residues are self-correlated in a higher order. The p-value in the second case, 0.14, is better than the one obtained in the first case: 0.02. However, the results are in the indecision zone and the test is not conclusive (Table 6).

The linearity of the model
Ramsey test: where H0 represents the hypothesis of a linear model, and H1: the model is not linear (Miles and Huberman, 1994). Statistics F is used to reject the null hypothesis. Thus, if the probability associated with the F test is lower than the 5% threshold, the null hypothesis that the model is linear is rejected. As the probability associated with the test is 0.94, greater than 5%, the null hypothesis according to which the model is linear up to the third-order power is accepted (Table 7).

First-order error autocorrelation. The Durbin-Watson Test
Hypothesis H0 assumes that errors are not self-correcting first. H1: errors are self-correcting. Considering that these statistics are tabulated, the values depend on the level of significance chosen (5%), the number of observations considered and the number of variables. Thus, it is necessary to determine the critical values d1 and d2. In this case, the table values for the DW test are as follows: d1 = 1.78469, d2 = 1.80075. Because DW = 1.7235 and is in the range [0; 1.78469], the null hypothesis is rejected, so there is first-order autocorrelation (Table 8).

CONCLUSION

This study shows how to use and interpret econometrics and data analysis techniques.
Statistics represent the art and science of collecting and understanding data that characterize mass phenomena.
This article has presented, summarized, applied and interpreted some of the basic techniques of statistics and econometrics, such as: descriptive analysis, distribution analysis, logistic regression, as well as the most frequently used tests for model validation. Its conclusions are as follows:
- there is a strong link between the profitability of the Apple asset and the profitability of the market portfolio;
- if the profitability of the market portfolio increases, the profitability of the Apple asset will increase;
- the series of the two indicators are stationary;
- there is a positive interdependence between the two returns;
- the variation of the market explains 33% of the variation of the asset according to the regression.

REFERENCES

[1] Adler, P. A. and Adler, P. (1994). Observational techniques. In *Handbook of Qualitative Research*, edited by N. K. Denzin and Y. S. Lincoln, pp. 377–392.
[2] Angrist, J.D. and J.S. Piche (2009). *Mostly Harmless Econometrics*, Princeton University Press.
[3] Harvey, A.C. (1991). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
[4] Miles, M. B. and Huberman, A. M. (1994). Qualitative Data Analysis: An Expanded Sourcebook, 2nd ed. Thousand Oaks, CA: SAGE.
[5] Sharpe, W.F. (1964). *Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*. Journal of Finance, 19.
[6] Wright, Sewell (1921). *Correlation and causation*, Journal of Agricultural Research, 20,557-585.
[7] Wu, De-Min (1973). *Alternative tests of independence between stochastic regressors anddisturbances*, Econometric, 41, 733-750.

Figures & Tables

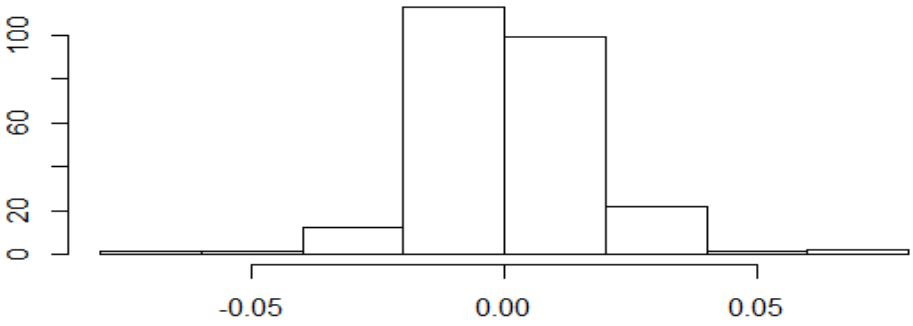

Figure No. 1
The histogram of the profitability of the asset
Source: Authors' own research results / contribution

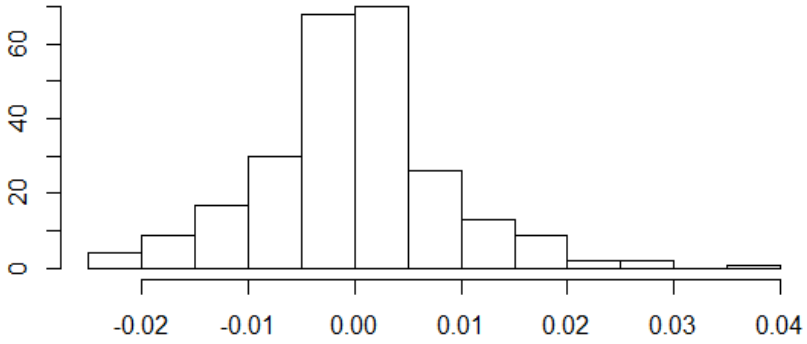

Figure No. 2
The histogram of the profitability of the market portfolio
Source: Authors' own research results / contribution

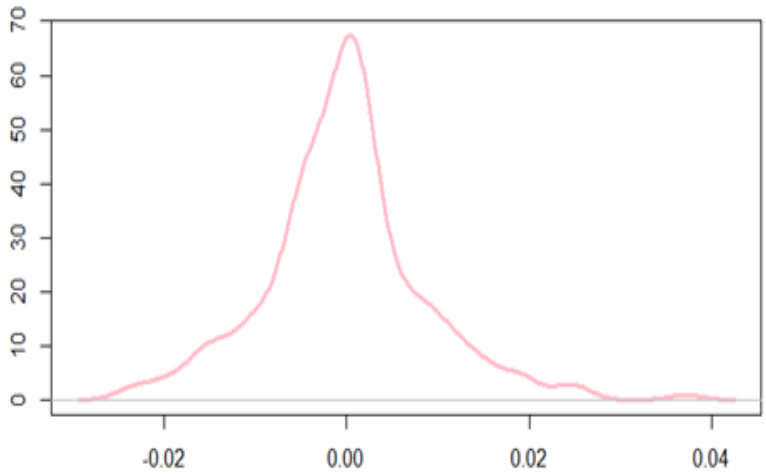

Figure No. 3
Probability densities of profitability of the market portfolio
Source: Authors' own research results / contribution

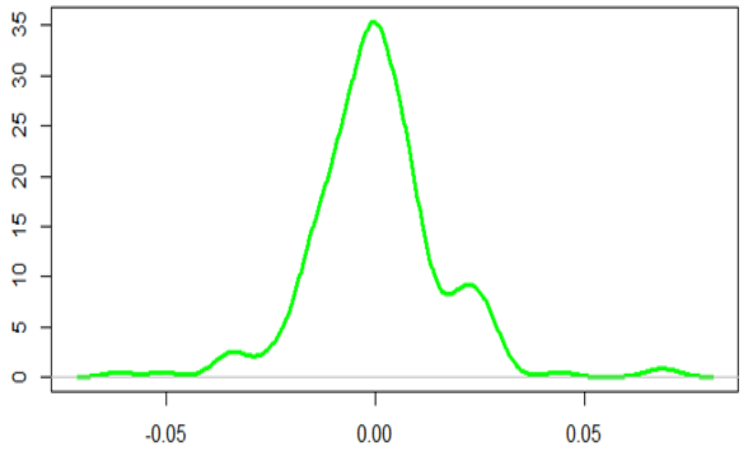

Figure No. 4
Probability densities of profitability of the asset
Source: Authors' own research results / contribution

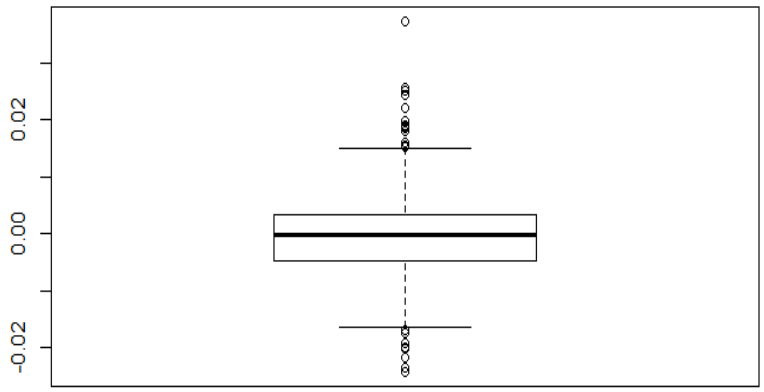

Figure No. 5
Boxplot for the profitability of the market portfolio
Source: Authors' own research results / contribution

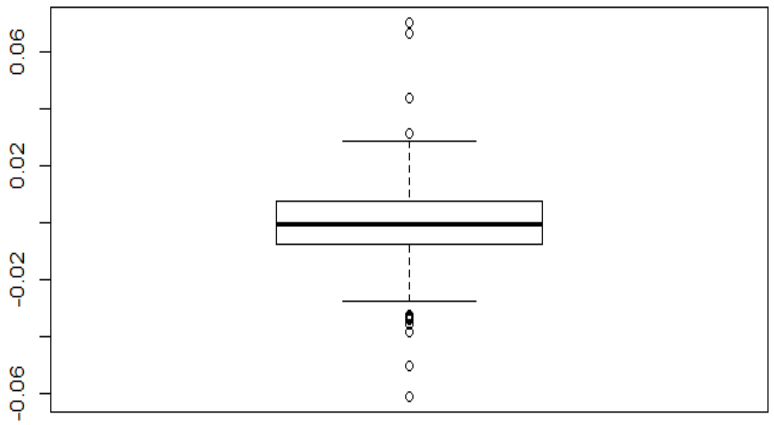

Figure No. 6
Boxplot for the profitability of the asset
Source: Authors' own research results / contribution

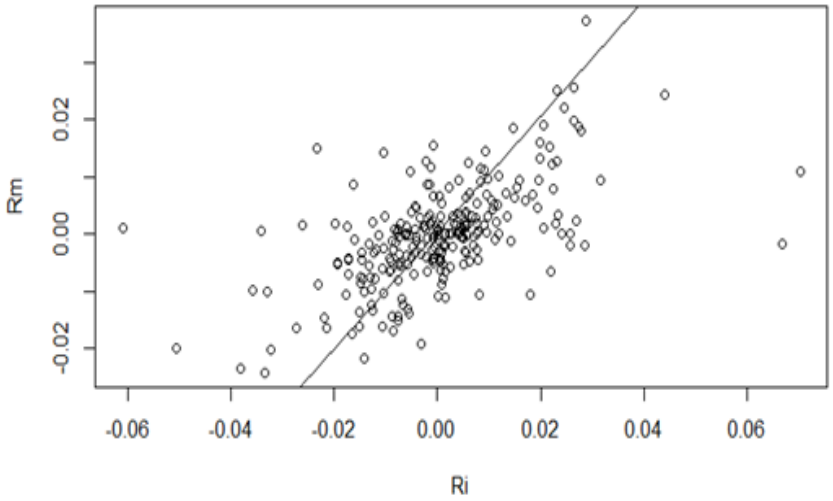

Figure No. 7
The interdependence between returns
Source: Authors' own research results / contribution

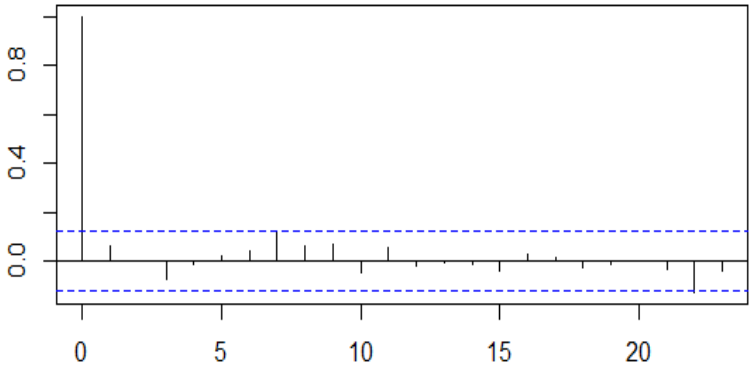

Figure No. 8
The correlation of the profitability of the asset
Source: Authors' own research results / contribution

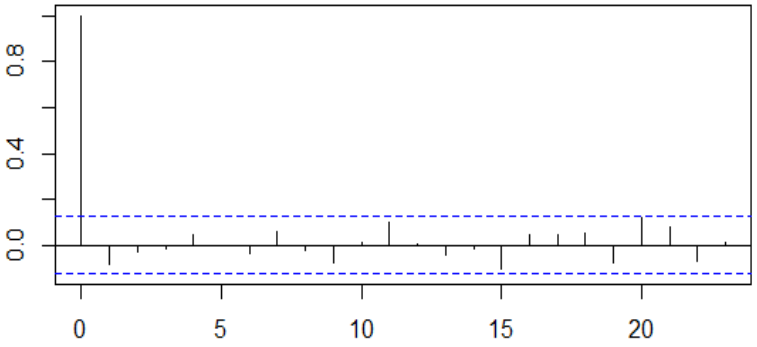

Figure No. 9
The correlation of the profitability of the market portfolio
Source: Authors' own research results / contribution

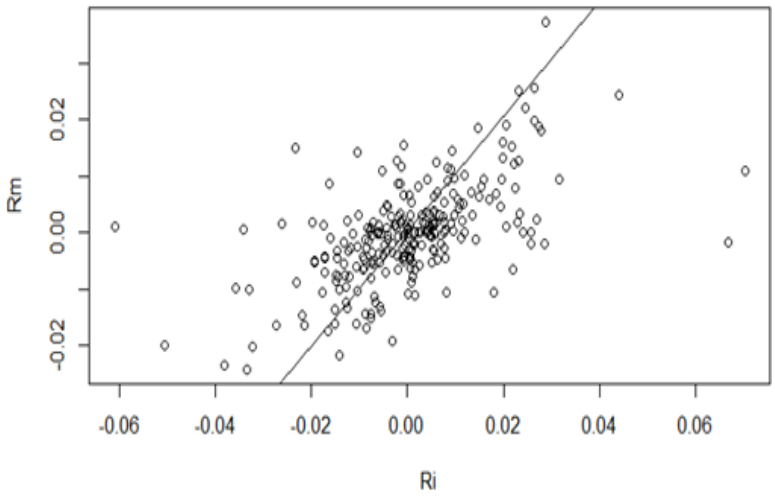

Figure No. 10
Interdependence between asset return and market portfolio return
Source: Authors' own research results / contribution

Table No. 1
Quantiles of market portfolio series (Rm) and Apple asset return series (Ri)

Quantile	Rm	Ri
0%	-0.02416	-0.061
25%	-0.00473	-0.00767
50%	-0.00016	-0.00044
75%	0.003266	0.007536
100%	0.037258	0.070328

Source: Authors' own research results / contribution

Table No. 2
Jarque-Bera test for normality of market portfolio series (Rm) and Apple asset return series (Ri)

data	Ri	Rm
JB	139.99	34.006
p-value	<2.2e-16	<2.2e-16

Source: Authors' own research results / contribution

Table No. 3
Stationarity test for market portfolio series (Rm) and Apple asset return series (Ri)

data	Ri	Rm
Dickey-Fuller	-5.2086	-5.7486
Lag order	6	6
p-value	0.01	0.01
alternative hypothesis	stationary	stationary

Source: Authors' own research results / contribution

Table No. 4
Regression function

Call:	lm(formula=Ri ~ Rm)			
Residuals:				
Min	1Q	Median	3Q	Max
-0.062649	-0.006243	-0.000039	0.005843	0.067996
Coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0004306	0.0007889	0.546	0.586
Rm	1.0147795	0.0894605	11.343	<2e-16
Residual standard error:	0.01249 on 249 degrees of freedom			
Multiple R-squared:	0.3407			
Adjusted R-squared:	0.338			
F-statistic:	128.7 on 1 and 249 DF			
p-value:	<2.2e-16			

Source: Authors' own research results / contribution

Table No. 5
Output of regression

	2.50%	97.50%
(Intercept)	-0.00112	0.001984
Rm	0.838584	1.190975

Source: Authors' own research results / contribution

Table No. 6
Homescedasticity test

Breusch-Godfrey test for serial correlation of order up to 1	
LM tests	4.7438
df	1
p-value	0.0294
Breusch-Godfrey test for serial correlation of order up to 3	
LM tests	5.3243
df	3
p-value	0.1495

Source: Authors' own research results / contribution

Table No. 7
Liniarity test

RESET	0.005366
df1	1
df2	2
p-value	0.9417

Source: Authors' own research results / contribution

Table No. 8
Autocorrelation of the errors test

DW	1.7235
p-value	0.01414
alternative hypothesis	true autocorrelation is greater than 0

Source: Authors' own research results / contribution